

Research

Open Access

RNase MRP and the RNA processing cascade in the eukaryotic ancestor

Michael D Woodhams^{†1}, Peter F Stadler², David Penny¹ and
Lesley J Collins^{*†1}

Address: ¹Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand and ²Bioinformatics Group, Department of Computer Science and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107, Germany

Email: Michael D Woodhams - M.D.Woodhams@massey.ac.nz; Peter F Stadler - Peter.Stadler@bioinf.uni-leipzig.de;
David Penny - D.Penny@massey.ac.nz; Lesley J Collins* - L.J.Collins@massey.ac.nz

* Corresponding author †Equal contributors

from First International Conference on Phylogenomics
Sainte-Adèle, Québec, Canada. 15–19 March, 2006

Published: 8 February 2007

BMC Evolutionary Biology 2007, 7(Suppl 1):S13 doi:10.1186/1471-2148-7-S1-S13

© 2007 Woodhams et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Within eukaryotes there is a complex cascade of RNA-based macromolecules that process other RNA molecules, especially mRNA, tRNA and rRNA. An example is RNase MRP processing ribosomal RNA (rRNA) in ribosome biogenesis. One hypothesis is that this complexity was present early in eukaryotic evolution; an alternative is that an initial simpler network later gained complexity by gene duplication in lineages that led to animals, fungi and plants. Recently there has been a rapid increase in support for the complexity-early theory because the vast majority of these RNA-processing reactions are found throughout eukaryotes, and thus were likely to be present in the last common ancestor of living eukaryotes, herein called the Eukaryotic Ancestor.

Results: We present an overview of the RNA processing cascade in the Eukaryotic Ancestor and investigate in particular, RNase MRP which was previously thought to have evolved later in eukaryotes due to its apparent limited distribution in fungi and animals and plants. Recent publications, as well as our own genomic searches, find previously unknown RNase MRP RNAs, indicating that RNase MRP has a wide distribution in eukaryotes. Combining secondary structure and promoter region analysis of RNAs for RNase MRP, along with analysis of the target substrate (rRNA), allows us to discuss this distribution in the light of eukaryotic evolution.

Conclusion: We conclude that RNase MRP can now be placed in the RNA-processing cascade of the Eukaryotic Ancestor, highlighting the complexity of RNA-processing in early eukaryotes. Promoter analyses of MRP-RNA suggest that regulation of the critical processes of rRNA cleavage can vary, showing that even these key cellular processes (for which we expect high conservation) show some species-specific variability. We present our consensus MRP-RNA secondary structure as a useful model for further searches.

Background

There is high interest in discovering new roles of RNA in modern eukaryotes [1-4]. The number of putative ncRNAs (non-coding RNAs) in the mammals alone has increased about 20-fold in the last five years [1], thus any information on the origins and functions of well-established ncRNAs is relevant and timely. In eukaryotes a number of ncRNA-based molecules are directly involved in the cleavage and processing of other RNA molecules. A classic example is the cleavage of rRNA transcript by RNase MRP, a ribonucleoprotein complex consisting of a single RNA molecule and about 10 proteins [5-7]. The processing of RNA by RNA can extend through several layers such as the snRNAs (small nuclear RNAs) in the spliceosome release snoRNAs (small nucleolar RNAs) from introns which in turn are involved in the modification of rRNA, tRNA or snoRNAs (see Figure 1). The network of these processes is called the eukaryotic RNA-processing cascade [8]. This cascade centres on the processing mRNA, tRNA and rRNA and although each of these RNAs is cleaved in separate reactions, there are linkages between these reactions as shown in Figure 1. The question we ask here is: how ancient are these RNA-based processes?

Pre-mRNA contains introns that are processed by the spliceosome (consisting of 5 snRNAs and ~200 proteins) [9-11] but there is also further processing such as the addition of the 5'-cap and 3' poly-A-tail [12]. Although the capping and polyadenylation processes are not RNA-based reactions they do include some proteins found in the spliceosome [9]. The snRNAs within the spliceosomal complex not only direct the binding and coordination of the splice sites but are also implicated in the catalysis of the splicing reactions [13]. Some introns contain within them ncRNAs such as snoRNAs (involved in modification of rRNA, tRNA and snRNAs, reviewed in [14]) or miRNAs involved in the degradation of mRNA [15-17]. Similarly, pre-tRNA is processed by RNase P; a ribonucleoprotein consisting in eukaryotes of a single RNA and about 8-10 proteins [18]. RNase P (abbreviated here as P) is found throughout eukaryotes and prokaryotes [18,19], and thus may date back to the RNA-world [20,21]. Pre-rRNA is heavily processed and the A3 site in the ITS region is cleaved by the ribonucleoprotein RNase MRP (abbreviated here as MRP) generating the mature 5.8S rRNA [22-25].

MRP (Mitochondrial RNA Processing) was originally identified as an RNA-protein endoribonuclease that processes RNA primers for DNA replication in the mitochondria [26]. However, the majority of MRP (99%) is observed in the nucleolus where it is involved in pre-rRNA processing [18]. MRP probably has other essential functions [27] including roles in chromosomal segregation [28] and control of cell division [29]. Initial evolutionary

studies (including [7]) only used MRPs from animals (13 mammals and frog), yeasts (20 Saccharomycetalian yeasts plus the fission yeast *S. pombe*) and plants (two dicotyledons). Although from the main multicellular groups, these sequences covered only a limited range within each group: land vertebrates within metazoans, Ascomycota within fungi, and the core eudicots within plants, leaving open the question whether MRP was present in the last common ancestor of eukaryotes [7].

We earlier [7] considered three hypotheses for the distribution of MRP (Figure 2). The first is that MRP is very ancient, occurring at least in the first eukaryotes. There are many variants on this model, and MRP could even be older in that most catalytic roles of RNA may derive from earlier stages in the origin of life, such as in the RNA-world [20]. Hypothesis B is that MRP arose from a duplication of P within modern eukaryotes (i.e. after the Eukaryotic Ancestor). This predicts a limited distribution of MRP in eukaryotes as well as explains the observation that P and MRP share most of their associated proteins [18]. Such duplication would be followed by specialization of the paralogous complexes, P being restricted to tRNA, and MRP to rRNA. There are several sub-hypotheses under this model, whether MRP took a new role in an internal excision in the rRNA precursor, or whether in eukaryotes, P initially carried out both reactions (with the precursors of tRNA and of rRNA) [7]. Hypothesis C is that MRP is derived from an early mitochondrial RNase P, followed by transfer of the gene to the nucleus, and co-option of MRP to a role in the nucleus in processing rRNA. Thus it is unclear whether the first role of MRP was in the nucleus and was later co-opted into a role in the mitochondrion, or vice versa.

In our earlier work it was concluded that the Hypothesis B was the most likely, that MRP had arisen within eukaryotes by a duplication of P, with subsequent specialization. The evidence against MRP coming from mitochondria (Hypothesis C) was that the secondary structure of MRP-RNA, as measured by RNA-shape metrics [7], was more similar to the eukaryote RNase P RNA than to the RNA from bacterial RNase P (the presumed source of the mitochondrial RNase P). Similarly, the apparent limited distribution of MRP in eukaryotes (at that time only in animals, fungi and plants) made it seem unlikely that MRP was present in the ancestral eukaryote. This left the duplication of P within eukaryotes (Hypothesis B) as the most likely. Two developments have changed our conclusions. Firstly, it now appears that the plant lineage and the fungi and animal lineage are widely separated on the eukaryote tree [30] and secondly, the recent characterisation of MRP in additional groups (as reported here and [31]) means that our initial conclusion must be reconsidered.

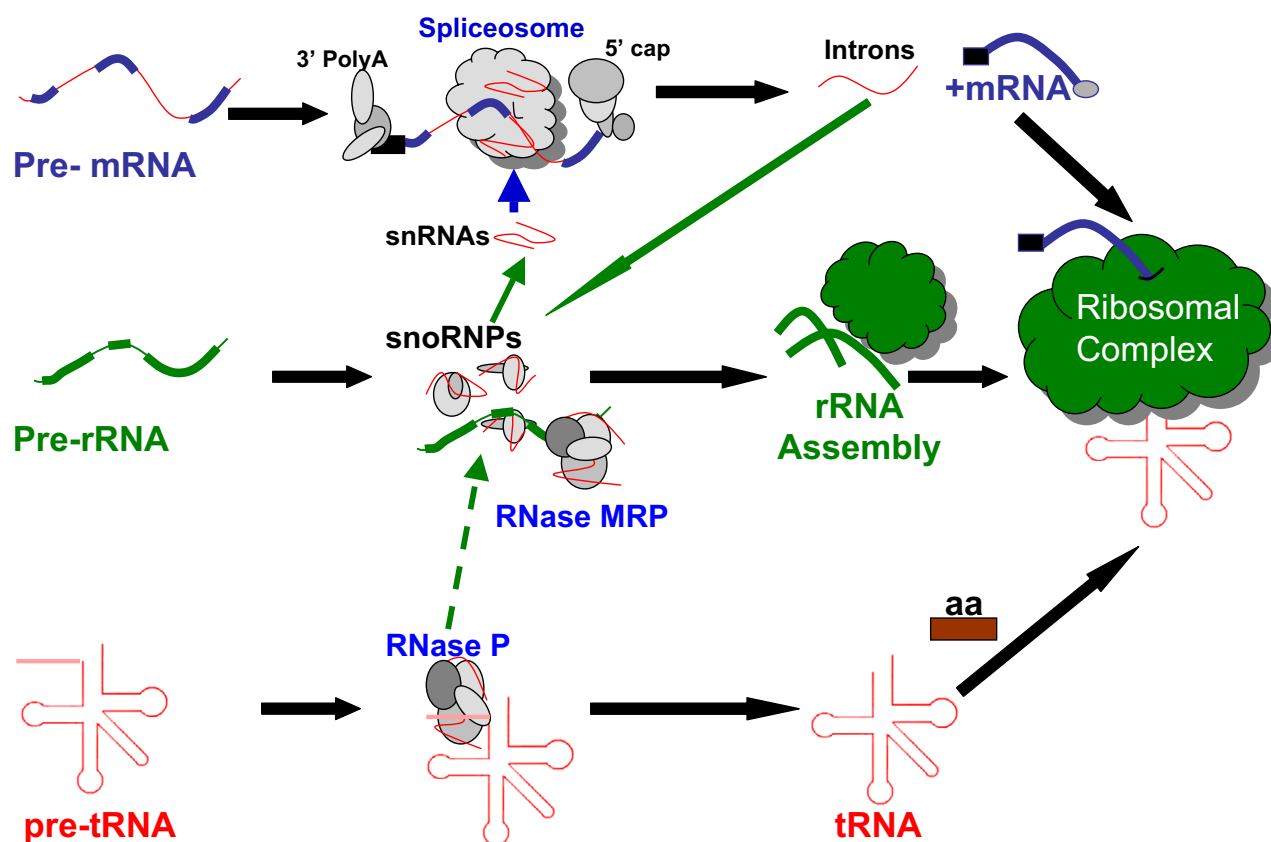


Figure 1

The eukaryotic RNA-processing cascade. mRNA is cleaved by the spliceosome (comprised of snRNAs and proteins) to release the processed mRNA and introns. Some introns contain snoRNAs which in turn modify snRNAs, tRNAs and rRNAs. RNase P (P) cleaves pre-tRNA while RNase MRP (MRP) cleaves rRNA. The ribosomal complex (comprised of rRNAs) brings the tRNAs and mature mRNAs together for translation. The involvement of RNase P in pre-rRNA processing has been questioned recently in [59].

Piccinelli [31] recently extended the range of species from which MRP-RNA is characterised to include several protist species including apicomplexa. We have also used an MRP-search strategy to find candidate MRP-RNA sequences in other eukaryotic species. We have further examined MRP-RNA secondary structure and promoter regions from all known sequences to strengthen consensus models for MRP-RNA throughout eukaryotes. In the light of these results we discuss the presence of MRP in the last common ancestor of modern eukaryotes and re-examine its evolution and its relationship within the RNA-processing cascade throughout eukaryotic lineages. Because the deep divergences of eukaryotes is not known [30,32], our strategy has been to find candidate RNAs in as many lineages as possible, thus making our conclusions independent of the precise rooting of the eukaryotic tree [33].

Results

RNase MRP found throughout eukaryotes using specific search strategies

MRP-RNA as a non-coding RNA (ncRNA) is not easy to identify in genomic sequences. Piccinelli [31] used a strategy based on hidden Markov Models (HMMs) of the P4 region of the MRP-RNA secondary structure to identify it in many eukaryotes. Our search strategy (in Materials and Methods) was also based on this P4 region and found MRP-RNA sequences from additional species. A eukaryotic tree (based on [32]) showing species from which MRP-RNA has been characterised is shown in Figure 3, and a full list of species is given in Additional File 1.

Our MRP-RNA candidates were checked against existing gene records and EST databases to support our bioinformatic approach. Of our new MRP-RNA candidates, two

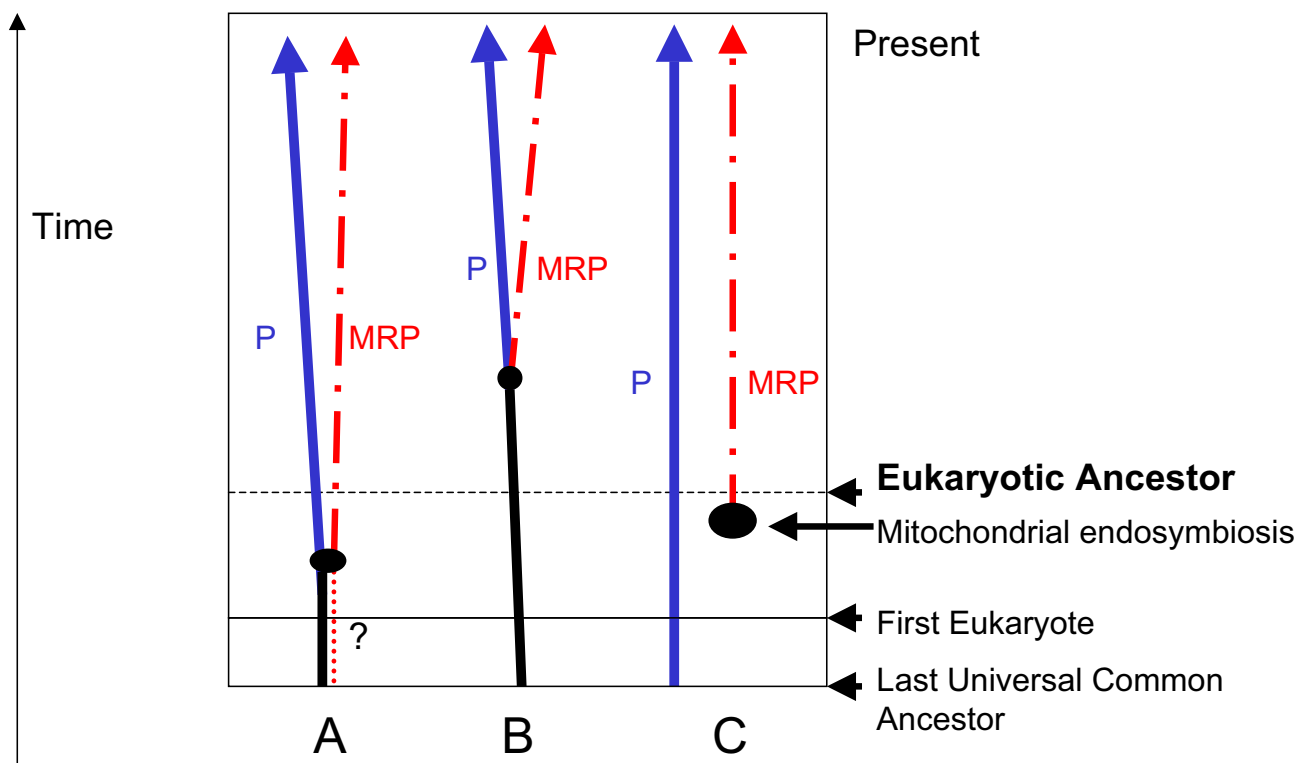


Figure 2

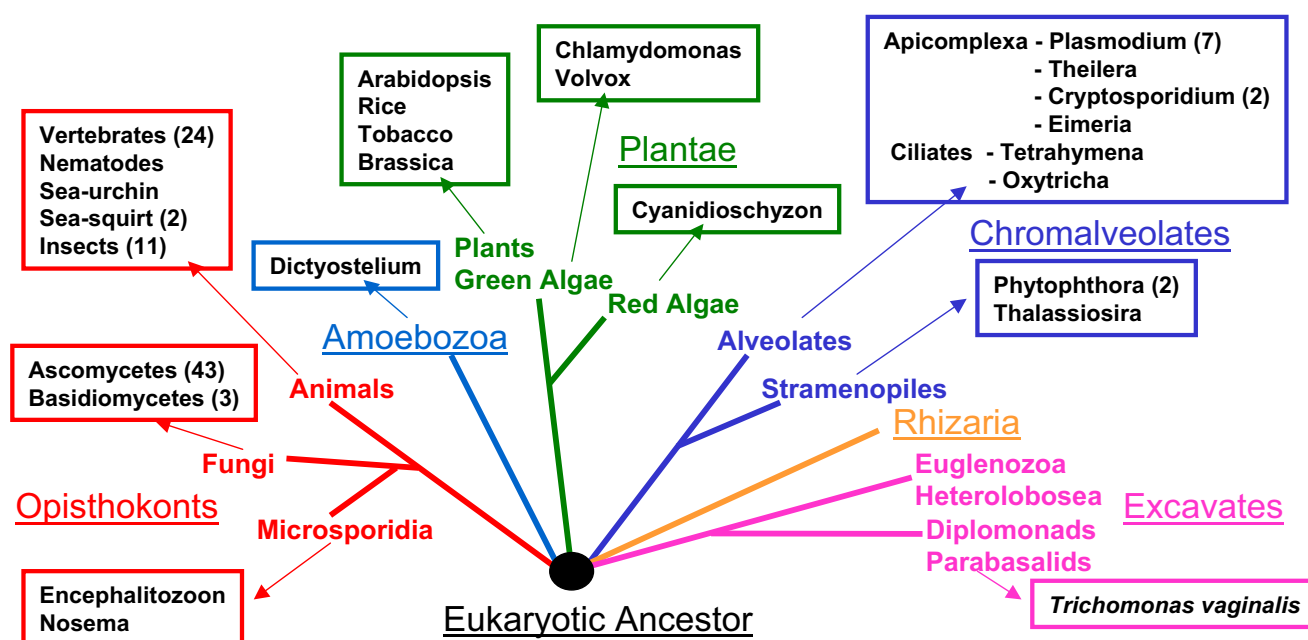
Hypothesis for the origin of MRP. This figure is based on [7], but modified to reflect present hypothesis of mitochondrial and eukaryotic evolution. The large black dots represent the point of duplication of the P-MRP ancestor. A: MRP was present in the last common ancestor of modern eukaryotes (the Eukaryotic Ancestor). Alternatively both MRP and P could have been present in the Last Universal Common Ancestor. B: MRP arose from a duplication of P after the Eukaryotic Ancestor, but before the ancestor of animals, fungi and plants. C: MRP arose from an early mitochondrial P within the Eukaryotic Ancestor.

are found in EST databases: *D. melanogaster* [GenBank: [CO153932](#)] and *Plasmodium yoelii* [GenBank: [BM161600](#) and GenBank: [BM160961](#)]. Validation of the MRP-RNA candidate from *Cryptosporidium parvum* with RT-PCR shows that this sequence is expressed in trophozoites (M. Irimia, data not shown). These *Plasmodium* and *C. parvum* results are particularly important because they show expression outside the animal, fungi and plant groups. In addition, five of our MRP-RNA sequences occur in Genbank records (excluding total sequence data): *G. gallus* [GenBank: [AADN01006913](#)], *T. nigroviridis* [GenBank: [CAAE01012081](#)], *P. falciparum* [GenBank: [NC_004325](#)] and *P. yoelii* [GenBank: [AABL01002665](#)]. That these do not overlap coding sequences is additional support for our computational approach. Interestingly, the candidate for the *D. melanogaster* MRP-RNA is on the negative strand of an intron in the Dmel_CG10365 gene [GenBank: [AE014297](#)], although the function of this gene is still unknown.

A single copy of the MRP-RNA gene was found in most species. However, the sea-urchin, *Strongylocentrotus purpur-*

ratus, appears to have five closely related sequences, although some could turn out to be artefacts of the current assembly. Multiple MRP-RNA genes have been observed in plants [34]. Humans [35] typically have a single true copy and a number of pseudogenes and studies of the MRP in the pufferfish *Takifugu rubripes* [36] indicate only a single copy in this species.

As before [7,31] we did not find any MRP-RNA candidate in the Diplomonad *Giardia lamblia*. We examined the rRNA organisation in *G. lamblia* to determine if an A3 site, normally cleaved by MRP, was present. The order of rRNA subunits are in general similar throughout eukaryotes [37] (Figure 4A), though there is variation in the length and composition of ITS regions. Some microsporidia (e.g. *E. cuniculi* and *Nosema apis*) contain no ITS2 region and have no cleavage between the 5.8S and the 28S rRNA subunits. Another group of microsporidia, including *Nosema bombycis* and *N. spodopterae*, are exceptions to the standard eukaryote ordering, having the fused 5.8S/28S subunit before the 18S rRNA subunit [38]. *G. lamblia* rRNA has the standard ordering of eukaryotic rRNA, and its ITS1 region

**Figure 3**

Distribution of MRP in eukaryotes. Only some sub-branches are given for each of the main groups. To date MRP has not yet been characterized from *Giardia* (Diplomonads), *Trypanosoma* (Euglenozoa), *Entamoeba* (Amoebozoa) and *Caenorhabditis* (Animals – nematodes). The general structure of the eukaryotic tree is based on [33].

can be folded into a secondary structure with a short six nucleotide single stranded region between the two helices, a possible A3 cleavage site (Figure 4B). The ITS1 region from *Trichomonas vaginalis* (from which an MRP has been characterised) folds into a helix followed by an AC rich single-stranded region which could be the cleavage site for MRP. Future experimental analysis will determine the exact cleavage sites in these ITS regions.

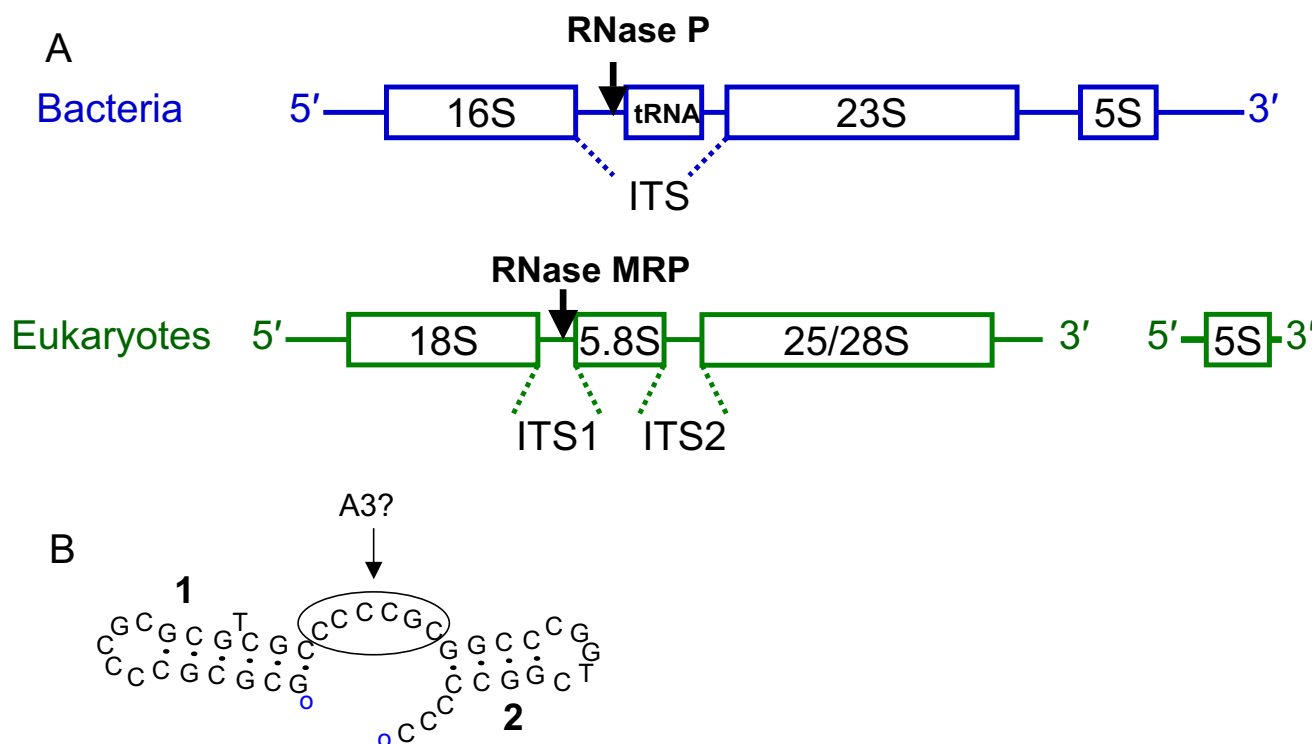
The characterisation of MRP across a wide range of eukaryotes indicates that the evolutionary relationship between MRP and P is ancient, and that both MRP and P are likely to have been present in the last common ancestor of modern eukaryotes. Although seemingly obvious, this distribution analysis importantly places MRP in the RNA-processing cascade present in the Eukaryotic Ancestor and further allows us to examine other characteristics of MRP and its relationship to other processes present in this ancient cascade.

Promoter analysis of candidate MRP-RNA sequences

With MRP the genes for proteins and for RNA are transcribed by different RNA polymerases; the protein genes by RNA Polymerase II, and the RNA by RNA Polymerase III (Type III promoter for U6 snRNA, 7SK and P-RNA; Type I for 5S rRNA.) [35,39,40]. RNA polymerase III transcription of MRP-RNA, P-RNA and U6 snRNAs has been

characterised for some animal, plant and fungal species. Analysis of upstream regions of MRP-RNAs and the literature allowed us to analyse the promoter elements associated with RNA polymerase III transcription of MRP-RNA to determine if there was any conservation of its promoter elements.

We find that MRP-RNA is probably transcribed by RNA polymerase III throughout eukaryotes but the set of RNA Polymerase III promoter elements may vary (see Figure 5). In general, vertebrate and plant MRP-RNA promoter regions contain an upstream TATA box, Proximal Sequence Element (PSE or USE) and a Distal Sequence Element (DSE) which can contain SP1, Staf and/or Octamer motifs [34,39-43]. In humans, the presence of the TATA box determines RNA polymerase specificity (i.e. RNA polymerase II or RNA polymerase III), with the other elements (e.g. PSE and DSE) enhancing transcription [44]. Plants require both the TATA box and the USE promoter (similar to the PSE element in vertebrates) with polymerase specificity determined by the spacing between the two elements [45]. In *Drosophila*, specificity is determined by the presence of the TATA box and the sequence of the PSE element [46]. However, the yeast *S. cerevisiae* uses a different RNA polymerase III promoter structure [44]. For example, the U6 snRNA promoter (similar to that expected in MRP-RNA) lacks PSE and DSE elements

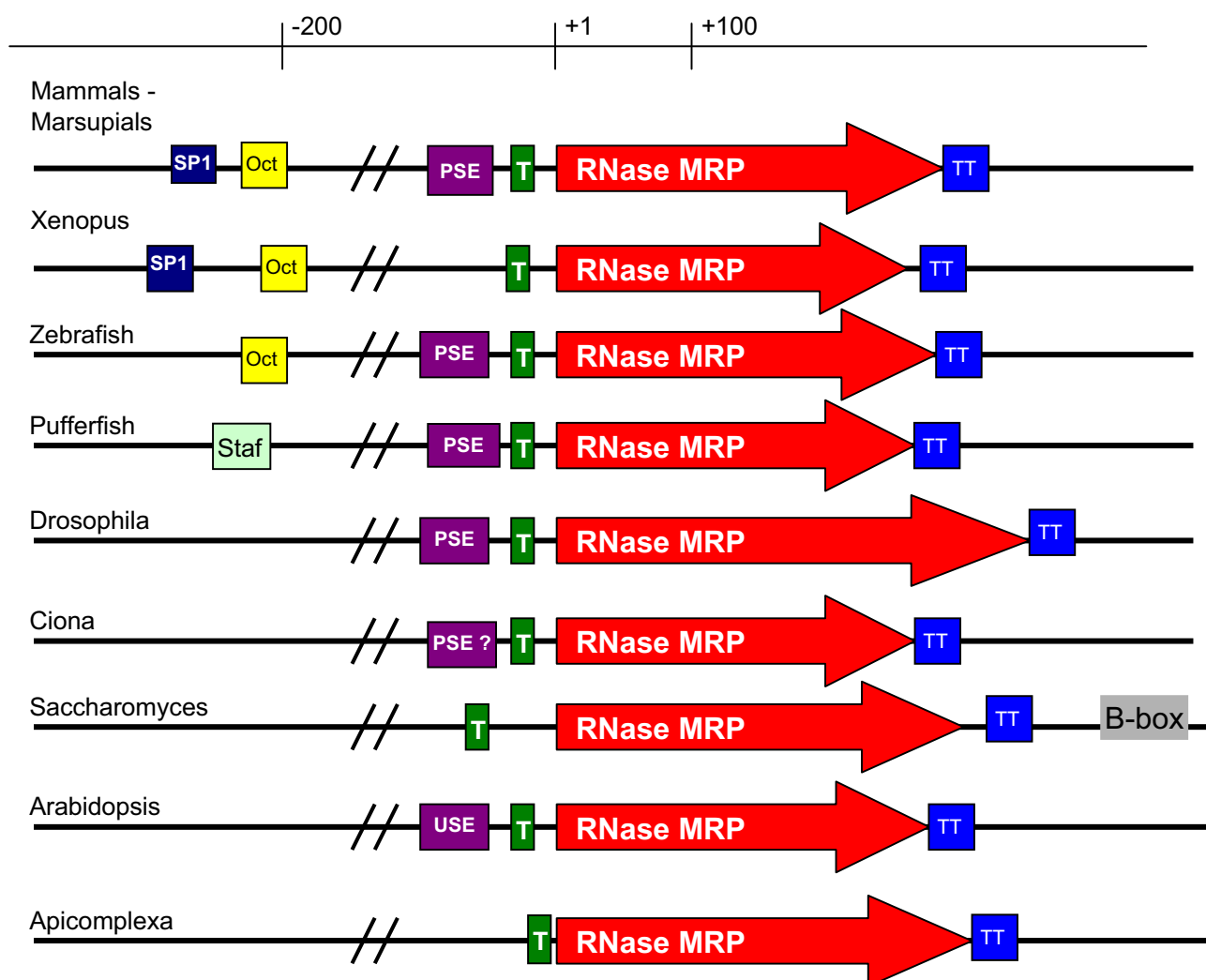
**Figure 4**

Promoter regions of Human MRP, P and U6. **A.** Organization of pre-rRNA based on [37]. In bacteria rRNA genes are co-transcribed as a polycistronic precursor (although exceptions are common). Most eukaryotes vary only in the length of their ITS regions, an extreme case being the microsporidian *Encephalitozoon cuniculi* which has completely lost its ITS2 having a fused 5.8S/28S subunit. RNase P and RNase MRP do not cleave the main transcripts but trim the ends of their respective substrates (the tRNA or 5.8S rRNA) after cleavage by other enzymes. In eukaryotes the 5S rRNA is transcribed separately by RNA polymerase III. **B.** The Diplomonad *Giardia lamblia* has the usual order of rRNA subunits with short ITS regions, however RNase MRP has not yet been characterized from this species. RNAstructure folding of *G. lamblia* ITS1 [56] showing a single stranded region between two stem loops that could possibly be an A3 site. Other foldings of this sequence and foldings of other sequences (DQ157272 and AF239841) produce just a single stem-loop.

but instead includes a downstream B box ~120 nucleotides beyond the terminator [44].

Promoter comparisons show there can be differences in MRP-RNA promoter elements between some closely related species. For example in fish, the MRP-RNA promoter region for *Takifugu* previously described in [36] characterises a Staf promoter element (a binding site for the Staf transcriptional activator protein) in the DSE. We are unable to find any Staf-binding sequence in the other two fish MRP-RNAs. The Zebrafish and *T. nigroviridis* MRP-RNAs have potential SP1 binding sites, but as with the *Takifugu*, no Octamer sites could be determined. Mammals and chicken MRP-RNAs contain a similar arrangement of their MRP-RNA promoter elements. The

frog MRP-RNA promoter regions have sequence motifs similar to mammals with a slightly different spacing between the elements within the DSE (the SP1 binding site is further upstream). Comparisons of six species of *Drosophila* (*D. melanogaster*, *D. pseudoobscura*, *D. yakuba*, *D. mojavensis*, *D. virilis* and *D. ananassae*) show a conserved PSE element (consensus sequence gcTTAaATTC-CCAAct) 23 nucleotides upstream of a TATA box (consensus sequence taaAa) which is about 16 nucleotides upstream of the transcription start site. The range of RNA polymerase III promoter structures and the lack of information about these promoter elements from protists make it difficult to identify promoter elements from protist MRP-RNA genes. Preliminary analyses of promoter regions from apicomplexa (*Plasmodium* and *Cryptosporid-*

**Figure 5**

MRP-RNA gene arrangement. Genes transcribed by RNA polymerase III (type III) usually contain a PSE (proximal sequence element) consisting of a TATA signal and PSE motif, and a DSE (distal sequence element) consisting of either a SP1, Oct or Staf binding site. Distances shown are approximate only. Key: T – TATA signal; PSE/USE – ; Oct – Octamer binding site; SP1 – SP1 binding site; Staf – Staf binding site. ? – Possible site. TT – Poly T termination signal. B-box – Downstream B-box motif.

ium) indicate likely TATA boxes, but at this stage we cannot predict the presence of PSE or DSE elements. The AT richness of this region makes promoter prediction difficult until such time as we have more experimental information about promoters in apicomplexa. A table of promoter elements is available from the corresponding author upon request.

Secondary structure analysis of MRP-RNA

Analysis of the secondary structure of MRP-RNA (Fig. 6 and [25,31,47,48]) shows that the overall secondary structure is conserved throughout eukaryotes (Figure 6).

Our naming of secondary structures in Figure 6 follows the convention that MRP-features are named after putatively homologous P-RNA features [25,49]. Features P1, P2, P3a, P4, P8 and P10 are found throughout all the MRP-RNA characterised to date while other features are nearly universal (P3b, P9 and eP19). A few features are observed in a limited phylogenetic range. Our consensus structure in Figure 6 is largely sequence independent, showing only the most conserved sequence motifs. This type of structure is essential for generating useful structure models for future computational searches for MRP-RNA.

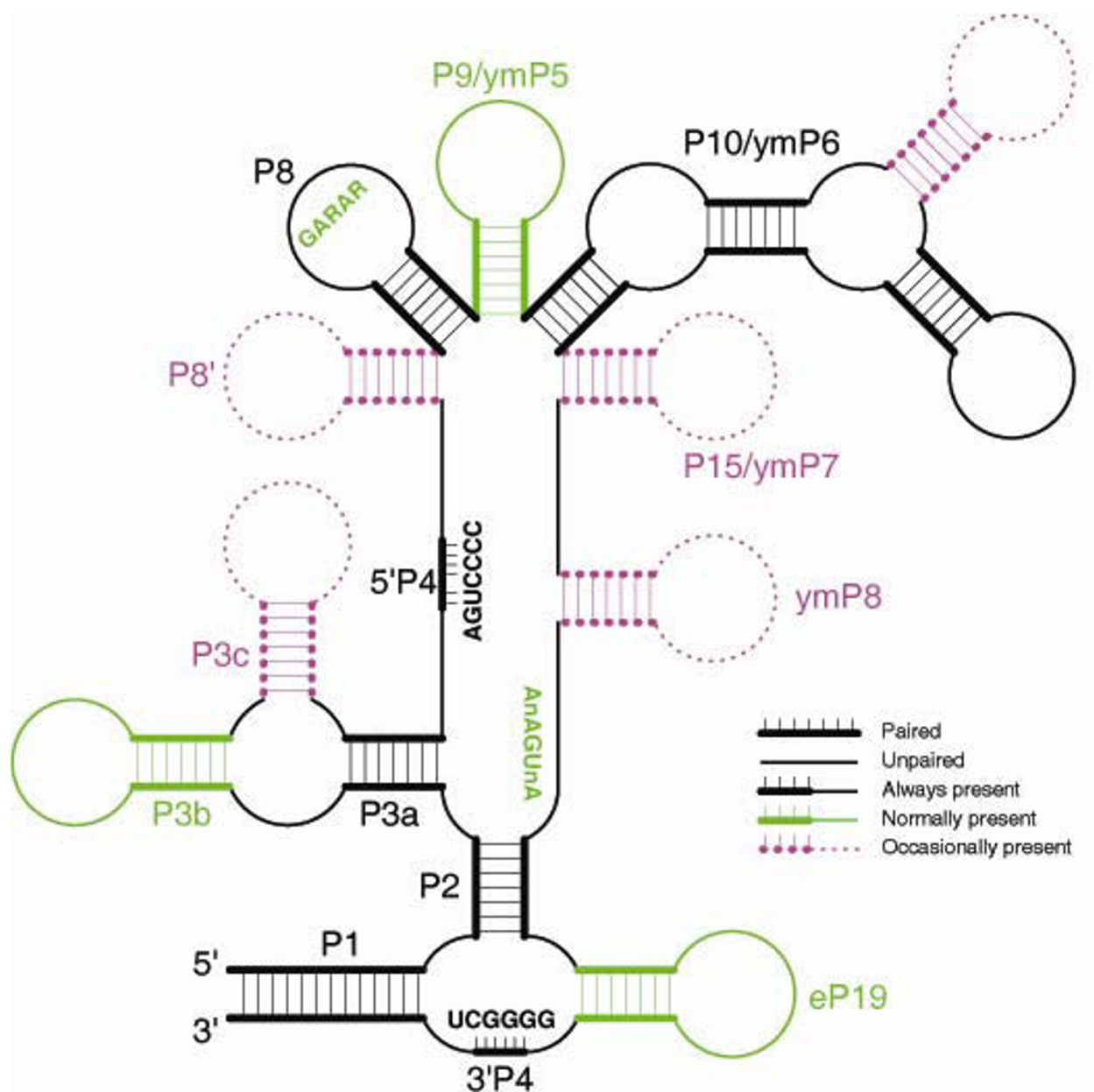


Figure 6

Summary diagram of MRP-RNA secondary structure. Black features (P1, P2, P3a, P4, P5, P7) are universally present. Green features are nearly universal, red features are observed in a few organisms of limited phylogenetic range. Thick lines are paired regions while unpaired regions are shown as thin lines. Conserved sequence motifs are indicated for the P4 (5' and 3') P8 and CRIV regions.

Details of the consensus MRP-RNA secondary structure are as follows. P3 nearly universally has two helices (P3a, P3b) separated by an internal loop; except for the absence of P3b in microsporidia and an additional P3c helix in Cryptosporidia, *Dictyostelium discoideum*, *Anopheles* and the nematode *Brugia malayi*. The P10 helix is often long,

typically 20–40 nucleotides, but in extreme cases (Pezizomycotina and Cryptosporidia) over 200 nucleotides. P15/ymP7 is present in Saccharomycetes yeasts and some apicomplexa. ymP8 is present in *S. pombe*, all Saccharomycetes with known MRP-RNAs, and some Pezizomycotina yeasts. The distinction between ymP7 and ymP8 is not

always clear (e.g. *Coccidioides immitis*). There are some species-specific divergences from our secondary structure model. P19 is absent from *Ciona intestinalis* and P6 is absent from microsporidia and *D. discoideum*. P6 is absent from Cryptosporidia in previously published secondary structures [31], but these sequences can have an alternative folding which includes the P6 helix.

The GARAR motif (R = A or G), recently noted by [31] and discussed by [25] is present in most species to date and is a defining feature of the P8 helix. It is usually in a pentaloop with an occasional deletion to a tetraloop (GARA) [25]. The major variants of the motif are GARAR and GARA but others are possible including the fish *Tetraodon nigroviridis* (CAAAG), cabbage *Brassica oleracea* (GAGG), *Babesia bovis* (TAAAG) and *Eimeria tenella* (GCGAG). Cryptosporidia, Plasmodia, *T. vaginalis* and some ascomycete fungi do not appear to have the GARAR motif. The three basidiomycete fungi also have varied GARAR motifs; two species (*Coprinus cinereus* and *Laccaria bicolor*) have GAAG as part of a bulge on P8. This region is suggested by [25] to be an MRP-specific region and thus will be important in the development of more MRP-specific search strategies. Another nearly universal motif is CR-IV (positioned 0–3 nucleotides before 3'P2). The sequence is AnAGUa, the 'U' and the first 'A' sometimes being substituted. This motif is not recognisable in *T. vaginalis* and some Alveolata. In *S. purpuratus* (sea-urchin) the motif is in a non-standard position, overlapping and extending beyond P2.

Discussion

The RNA cascade connects a number of RNA-based complexes where RNA is processing other RNA molecules. Figure 1 is a simplified model that shows the key processes and the main connections. Key processing complexes such as the spliceosome (snRNAs) [33], RNase P [19] and snoRNPs [50] are all seen as being ancestral to modern eukaryotes, even though details such as the intron recognition by the spliceosome [51] cannot yet be determined. The discovery of MRP across so many eukaryotes indicates that it was also part of this ancestral RNA-processing cascade. Given that MRP occurred so early in eukaryotes it is not surprising that MRP is now implicated in a number of other cellular processes (especially in well-studied species such as humans and yeast *S. cerevisiae*). As well as nuclear rRNA and mitochondrial primer cleavage functions, in *S. cerevisiae* at least, it has an additional function of promoting cell cycle progressing by cleaving CLB2 mRNA in its 5' UTR region at the end of mitosis [52,53]. It is possible that other functions of MRP may be found, especially when other RNA-processing systems are investigated.

One main conclusion from this study is that, with the placement of MRP in the RNA-processing cascade of the

Eukaryotic Ancestor, we see little change in basic RNA-processing throughout eukaryotes. Eukaryotes and prokaryotes have basic differences in their processing of their rRNA transcripts [37]; the main eukaryotic transcript contains ITS1 (between the 18S and 5.8S) and ITS2 (between the 5.8S and 28S) whereas prokaryotes generally have only an ITS1 with the 5' end of the prokaryotic 23S with strong homology to the eukaryotic 5.8S sequence [54] (Figure 4A). Thus we find the 5.8S rRNA, either cleaved as a separate subunit, or fused to the large rRNA subunit (no ITS2 present). Typically within eukaryotes we find the 5.8S rRNA cleaved but not in prokaryotes. There are exceptions for both, microsporidia do not appear to have an ITS2 [55,56], and in prokaryotes RNase III cleaved IVS (intervening sequence) regions in α -proteobacteria have been found [54]. RNase III, which is involved in cleaving the prokaryotic rRNA transcript has now been implicated in ITS1 processing in *S. pombe* [57]. It is likely that the Eukaryotic Ancestor contained cleaved 5.8S rRNA, but we cannot yet determine if the last universal common ancestor (of eukaryotes and prokaryotes) contained a separate or fused 5.8S.

The cleavage of site A3 in the ITS1 region by MRP is similar to the cleavage of the tRNA in the bacterial system by P. However, we do not know whether the eukaryotic-type of RNA-processing cascade has evolved from the bacterial RNA-processing system. Bacteria, archaea and eukaryotes have probably changed their original RNA-processing cascade, each in their own evolutionary trajectory [58]. Nor do we know whether the ancestral P had either a simplified bacterial-like form (with a single protein), or had multiple proteins (like eukaryotes) which have been reduced in prokaryotes to a single protein. One piece of evidence for this second hypothesis is that the human Rpp29 (pop4) protein, shared by P and MRP, acts as a cofactor for the *E. coli* P-RNA [25]. We can no longer make the assumption that the prokaryotic models of RNase P and RNA-processing are ancestors of the eukaryotic complexes.

The high similarity of secondary structure between MRP and P [7] is indicative of an evolutionary relationship, probably maintained by the sharing of numerous proteins between the MRP and P. However, it appears likely that P and MRP were already separate in the Eukaryotic Ancestor. The association of proteins with their respective MRP and P-RNAs may differ not only between P and MRP but may also vary between species [59]. Thus much of the large similarity in secondary structure between sections of MRP and P-RNAs (e.g. the P3-region indicated in [31]) is likely due to the constraints placed on the RNA molecules by their interactions with their common proteins even if some proteins interact transiently in some species.

The GARAG motif is an interesting addition to the MRP-RNA secondary structure. This pentaloop (or sometimes tetraloop) is potentially a protein or RNA binding site, which would explain its conserved nature. GNRA tetraloop motifs are also found in bacterial P (Type B but not Type B), archaeal P (both Type A and Type M) and also possibly in the ep9 helix of P from the yeast *S. cerevisiae* [25]. They are also common features of other ncRNAs. Identifying binding target sites for this motif in MRP may aid in understanding some of the differential protein binding [6]. The new consensus secondary structure model gives information required for future search strategies. Computational analysis of ncRNAs often use secondary structure and highly conserved sequence motifs rather than complete ncRNA sequences [7]. This model should allow more sophisticated MRP-RNA search algorithms to identify of MRP-RNA in additional eukaryotes.

The promoter region analysis indicates that RNA polymerase III is used throughout eukaryotes to transcribe MRP-RNA. It is interesting that there is such a range of promoter elements for MRP-RNA transcription and that the spacing between different elements appears critical in some species and not others. This may indicate that the regulation of MRP differs between groups of eukaryotes even for such an essential function as rRNA cleavage. There is little information about RNA polymerase III transcription in protists and the possible promoter regions are seldom reported when new ncRNAs are characterised.

Although MRP has been characterised in many eukaryotes, it has yet to be found in the nematodes *C. elegans* and *C. briggsae*, although complete genomes have been available for these species for several years. Similar exceptions are the protists *Giardia lamblia* and *Entamoeba histolytica*. However, MRP-RNA [31], was found in another nematode species *Brugia malayi* [31]. P-RNA has only recently been published for *C. elegans* [60] and *G. lamblia* [31,61] but even with all this new information we were still unable to find MRP-RNA in the above species. A recent survey for structured ncRNAs [62] based on comparative analysis of *C. elegans* and *C. briggsae* again did not result in a plausible MRP-RNA candidate. We have also not yet recovered any MRP-RNA from a *G. lamblia* RNA library (although we have recovered P-RNA) (S.X. Chen, data not shown). Nevertheless, in *C. elegans*, *G. lamblia* and *E. histolytica* the rRNA gene arrangement is generally the same as in other eukaryotes [63], although the ITS1 regions are very short in *G. lamblia* and *E. histolytica* [56]. Short ITS1 regions are also found in other species with reduced genomes such as *Trichomonas vaginalis* and the microsporidian *E. cuniculi* both of which have had MRP-rRNAs [31]. *G. lamblia* does not contain a nucleolus, but it is expected to contain a eukaryotic-like rRNA processing system due to the presence of many pre-rRNA processing proteins [64]. Some

proteins that are usually shared between P and MRP have also been found in *G. lamblia* [65]. Two MRP-specific proteins (proteins not found in P, Smn1 and Rmp1) [66] have been characterised only in yeast and thus using MRP-specific proteins to indicate the presence of MRP, is not an option at this stage. The large evolutionary distance between Diplomonads and the only excavate from which MRP has previously been characterised (the Parabasalid, *Trichomonas vaginalis* [31]) means that MRP may be difficult to characterise in *G. lamblia* and even other excavates such as the Euglenozoa (e.g. *Trypanosoma brucei*), using present techniques in computational analysis which rely heavily on sequence homology in the P4 region.

It is likely that the major protein and RNA components of the RNA processing cascade evolved before the Eukaryotic Ancestor (which is now seen to have come after the mitochondrial endosymbiosis [67]) It is interesting that MRP is still found in species that no longer contain mitochondria [31], but contain instead reduced organelles such as mitosomes or remnant mitochondria (apicomplexa and microsporidia) and hydrogenosomes (ciliates, parabasalids and some fungi) [67].

The RNA-processing cascade is now seen as a complex feature of the ancestral eukaryotic cell. Only when we understand which eukaryotic processes were in the Eukaryotic Ancestor we can then consider how they evolved.

Conclusion

We present the organisation of RNA-processing in eukaryotes as a cascade of RNA-based processing reactions cleaving or modifying other RNA molecules. The main components of this cascade are conserved throughout eukaryotes and are likely to have been present in the Eukaryotic ancestor. We can now place MRP in this cascade and thus basic RNA-processing has been preserved in eukaryotes. Analysis of MRP-RNA promoter regions suggest, however, that regulation of these critical processes differs between species showing that even these key cellular processes are showing some species-specific variability. Computational searches for ncRNAs are difficult due to the necessary incorporation of secondary structure as well as sequence information. Our consensus secondary structure for MRP-RNA provides a useful model for further search strategies.

Methods

Searching genomes for RNase MRP RNA

The conserved regions around the P4 pseudoknot are important for finding potential candidate MRP-RNAs. A genome is scanned for regions similar to the conserved sequences from known MRP-RNAs then candidates are evaluated for stereotypical secondary structure. Candidates with suitable secondary structure are evaluated for

upstream promoter regions expected for a gene transcribed by RNA polymerase III, then blasted against EST databases [68] for any indication that the candidate is expressed.

In more detail, the algorithm allows the search of genome sequences for closely located sequences which match or nearly match the consensus 5'P4 and 3'P4 regions (taken to be gaaAGuCCCC and acnnnanGGGGCUnannnu respectively, paired bases in uppercase). We have three levels of search criteria. Using the tightest criteria, 5'P4 and 3'P4 can be separated by 120 to 280 bases and we allow just one deviation from the consensus (either a single substitution of an unpaired base, or a substitution of a Watson-Crick pair by another such pair). The second, slightly looser, search criteria allows 100 to 360 bases between 5'P4 and 3'P4, and two deviations from the consensus. The loosest criteria allows a 80 to 500 base separation of P4, and two deviations from the consensus or a single violation of the expected Watson-Crick pairings. Genomes are scanned with the tightest criteria first, then with subsequent relaxed criteria if the first scan fails to identify an MRP candidate with suitable secondary structure. The Perl programs used are available upon request.

Secondary structure analysis of MRP-RNA

Vertebrate [47] and yeast [39] secondary structures were obtained from the literature. For each new candidate sequence, we use RNAfold [69] and Mfold [70] to fold sequences of varying lengths prior to the 5'P4 region, to find a candidate P3 structure, and similarly the region prior to 3'P4 to find a candidate P9 structure. If successful, this identifies a small region to search for the P2 structure. Once these three structures are identified, the complete structure is easily obtained. If the number of candidates from the scanning stage is large, we use RNAmotif [61,71] to filter out candidates that do not have suitable P2 and P3 helices (RNAmotif descriptor files are available on request). Where we have MRP candidates from closely related species, we refine our structures by comparing sequences with ClustalX [72] and DIALIGN [73], and comparing structures using a range of RNA comparison software (Alifold from the Vienna RNA package [74], RNAforester [75], RNASHapes [76] and RNACast [77]).

Authors' contributions

MW carried out the search and secondary structure analysis and drafted the original manuscript. PFS contributed to the search of new and not easily available genomes. DP participated in the design of the study and contributed to the evolutionary discussions. LC carried out the promoter analysis and drafted the final manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

MRP-RNA sequences from a wide distribution of eukaryotes.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S13-S1.xls>]

Acknowledgements

Thanks to Manuel Irimia for RT-PCR experiments, and Sylvia (Xiaowei) Chen (Allan Wilson Centre) for results from the *Giardia lamblia* RNA library. Computational analysis was carried out using the Helix Parallel Processing Facility at Massey University. This work was funded by the New Zealand Marsden Fund, the New Zealand Centres of Research Excellence Fund and the Bioinformatics Initiative of the German DFG.

This article has been published as part of *BMC Evolutionary Biology* Volume 7 Supplement 1, 2007: First International Conference on Phylogenomics. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcevolbiol/7?issue=S1>.

References

1. Huttenhofer A, Schattner P, Polacek N: **Non-coding RNAs: hope or hype?** *Trends Genet* 2005, **21**(5):289-297.
2. Mattick JS: **The functional genomics of noncoding RNA.** *Science* 2005, **309**(5740):1527-1528.
3. Washietl S, Hofacker IL, Lukasser M, Huttenhofer A, Stadler PF: **Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome.** *Nat Biotechnol* 2005, **23**(11):1383-1390.
4. Costa FF: **Non-coding RNAs: new players in eukaryotic biology.** *Gene* 2005, **357**(2):83-94.
5. van Eenennaam H, Jarrous N, van Venrooij WJ, Puijij GJ: **Architecture and function of the human endonucleases RNase P and RNase MRP.** *IUBMB Life* 2000, **49**(4):265-272.
6. Welting TJ, van Venrooij WJ, Puijij GJ: **Mutual interactions between subunits of the human RNase MRP ribonucleoprotein complex.** *Nucleic Acids Res* 2004, **32**(7):2138-2146.
7. Collins LJ, Moulton V, Penny D: **Use of RNA secondary structure for studying the evolution of RNase P and RNase MRP.** *J Mol Evol* 2000, **51**(3):194-204.
8. Poole A, Jeffares D, Penny D: **Early evolution: prokaryotes, the new kids on the block.** *Bioessays* 1999, **21**(10):880-889.
9. Jurica MS, Moore MJ: **Pre-mRNA splicing: awash in a sea of proteins.** *Mol Cell* 2003, **12**(1):5-14.
10. Kaufer NF, Potashkin J: **Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals.** *Nucleic Acids Res* 2000, **28**(16):3003-3010.
11. Lorkovic ZJ, Wiczeorek Kirk DA, Lambermon MH, Filipowicz W: **Pre-mRNA splicing in higher plants.** *Trends Plant Sci* 2000, **5**(4):160-167.
12. Shatkin AJ, Manley JL: **The ends of the affair: capping and polyadenylation.** *Nat Struct Biol* 2000, **7**(10):838-842.
13. Valadkhan S: **snRNAs as the catalysts of pre-mRNA splicing.** *Curr Opin Chem Biol* 2005, **9**(6):603-608.
14. Kiss T: **Biogenesis of small nuclear RNPs.** *J Cell Sci* 2004, **117**(Pt 25):5949-5951.
15. Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA: **Structural basis for double-stranded RNA processing by Dicer.** *Science* 2006, **311**(5758):195-198.
16. Ullu E, Tschudi C, Chakraborty T: **RNA interference in protozoan parasites.** *Cell Microbiol* 2004, **6**(6):509-519.
17. Ying SY, Lin SL: **Intronic microRNAs.** *Biochem Biophys Res Commun* 2005, **326**(3):515-520.
18. Xiao S, Scott F, Fierke CA, Engelke DR: **Eukaryotic ribonuclease P: a plurality of ribonucleoprotein enzymes.** *Annu Rev Biochem* 2002, **71**:165-189.

19. Hartmann E, Hartmann RK: **The enigma of ribonuclease P evolution.** *Trends Genet* 2003, **19**(10):561-569.
20. Jeffares DC, Poole AM, Penny D: **Relics from the RNA world.** *J Mol Evol* 1998, **46**(1):18-36.
21. Penny D: **An interpretive review of the origin of life research.** *Biol Philos* 2005, **20**(4):633-671.
22. Schmitt ME, Clayton DA: **Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in *Saccharomyces cerevisiae*.** *Mol Cell Biol* 1993, **13**(12):7935-7941.
23. Lygerou Z, Allmang C, Tollervey D, Seraphin B: **Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro.** *Science* 1996, **272**(5259):268-270.
24. Chu S, Archer RH, Zengel JM, Lindahl L: **The RNA of RNase MRP is required for normal processing of ribosomal RNA.** *Proc Natl Acad Sci USA* 1994, **91**(2):659-663.
25. Walker SC, Engelke DR: **Ribonuclease P: The evolution of an ancient RNA enzyme.** *Crit Rev Biochem Mol Biol* 2006, **41**(2):77-102.
26. Gold HA, Topper JN, Clayton DA, Craft J: **The RNA processing enzyme RNase MRP is identical to the Th RNP and related to RNase P.** *Science* 1989, **245**(4924):1377-1380.
27. Li K, Smagula CS, Parsons WJ, Richardson JA, Gonzalez M, Hagler HK, Williams RS: **Subcellular partitioning of MRP RNA assessed by ultrastructural and biochemical analysis.** *J Cell Biol* 1994, **124**(6):871-882.
28. Cai T, Reilly TR, Cerio M, Schmitt ME: **Mutagenesis of SNM1, which encodes a protein component of the yeast RNase MRP, reveals a role for this ribonucleoprotein endoribonuclease in plasmid segregation.** *Mol Cell Biol* 1999, **19**(11):7857-7869.
29. Clayton DA: **A big development for a small RNA.** *Nature* 2001, **410**(6824):29, 31.
30. Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW: **The Tree of Eukaryotes.** *Trends Ecol Evol* 2005, **20**(12):670-676.
31. Piccinelli P, Rosenblad MA, Samuelsson T: **Identification and analysis of ribonuclease P and MRP RNA in a broad range of eukaryotes.** *Nucleic Acids Res* 2005, **33**(14):4485-4495.
32. Simpson AG, Roger AJ: **The real 'kingdoms' of eukaryotes.** *Curr Biol* 2004, **14**(17):R693-696.
33. Collins L, Penny D: **Complex spliceosomal organization ancestral to extant eukaryotes.** *Mol Biol Evol* 2005, **22**(4):1053-1066.
34. Kiss T, Marshallsay C, Filipowicz W: **7-2/MRP RNAs in plant and mammalian cells: association with higher order structures in the nucleolus.** *Embo J* 1992, **11**(10):3737-3746.
35. Yuan Y, Reddy R: **5' flanking sequences of human MRP/7-2 RNA gene are required and sufficient for the transcription by RNA polymerase III.** *Biochim Biophys Acta* 1991, **1089**(1):33-39.
36. Myslinski E, Krol A, Carbon P: **Characterization of snRNA and snRNA-type genes in the pufferfish *Fugu rubripes*.** *Gene* 2004, **330**:149-158.
37. Lafontaine DL, Tollervey D: **The function and synthesis of ribosomes.** *Nat Rev Mol Cell Biol* 2001, **2**(7):514-520.
38. Tsai SJ, Huang WF, Wang CH: **Complete sequence and gene organization of the *Nosema spodopterae* rRNA gene.** *J Eukaryot Microbiol* 2005, **52**(1):52-54.
39. Li X, Frank DN, Pace N, Zengel JM, Lindahl L: **Phylogenetic analysis of the structure of RNase MRP RNA in yeasts.** *RNA* 2002, **8**(6):740-751.
40. Paule MR, White RJ: **Survey and summary: transcription by RNA polymerases I and III.** *Nucleic Acids Res* 2000, **28**(6):1283-1298.
41. Schmitt ME, Clayton DA: **Yeast site-specific ribonucleoprotein endoribonuclease MRP contains an RNA component homologous to mammalian RNase MRP RNA and essential for cell viability.** *Genes Dev* 1992, **6**(10):1975-1985.
42. Dermitzakis ET, Bergman CM, Clark AG: **Tracing the evolutionary history of *Drosophila* regulatory regions with models that identify transcription factor binding sites.** *Mol Biol Evol* 2003, **20**(5):703-714.
43. Topper JN, Clayton DA: **Secondary structure of the RNA component of a nuclear/mitochondrial ribonucleoprotein.** *J Biol Chem* 1990, **265**(22):13254-13262.
44. Huang Y, Maraia RJ: **Comparison of the RNA polymerase III transcription machinery in *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* and human.** *Nucleic Acids Res* 2001, **29**(13):2675-2690.
45. Waibel F, Filipowicz W: **The spacing between two promoter elements determines RNA polymerase specificity during transcription of U small nuclear RNA genes of *Arabidopsis*.** *Mol Biol Rep* 1990, **14**(2-3):149.
46. Li C, Harding GA, Parise J, McNamara-Schroeder KJ, Stumph WE: **Architectural arrangement of cloned proximal sequence element-binding protein subunits on *Drosophila* U1 and U6 snRNA gene promoters.** *Mol Cell Biol* 2004, **24**(5):1897-1906.
47. Frank DN, Adamidi C, Ehringer MA, Pitulle C, Pace NR: **Phylogenetic-comparative analysis of the eukaryal ribonuclease P RNA.** *RNA* 2000, **6**(12):1895-1904.
48. Walker SC, Aspinall TV, Gordon JM, Avis JM: **Probing the structure of *Saccharomyces cerevisiae* RNase MRP.** *Biochem Soc Trans* 2005, **33**(Pt 3):479-481.
49. Zhu Y, Stribinskis V, Ramos KS, Li Y: **Sequence analysis of RNase MRP RNA reveals its origination from eukaryotic RNase P RNA.** *RNA* 2006, **12**(5):699-706.
50. Yang CY, Zhou H, Luo J, Qu LH: **Identification of 20 snoRNA-like RNAs from the primitive eukaryote, *Giardia lamblia*.** *Biochem Biophys Res Commun* 2005, **328**(4):1224-1231.
51. Collins L, Penny D: **Investigating the Intron Recognition Mechanism in Eukaryotes.** *Mol Biol Evol* 2006, **23**(5):901-910.
52. Cai T, Aulds J, Gill T, Cerio M, Schmitt ME: **The *Saccharomyces cerevisiae* RNase mitochondrial RNA processing is critical for cell cycle progression at the end of mitosis.** *Genetics* 2002, **161**(3):1029-1042.
53. Gill T, Cai T, Aulds J, Wierzbicki S, Schmitt ME: **RNase MRP cleaves the CLB2 mRNA to promote cell cycle progression: novel method of mRNA degradation.** *Mol Cell Biol* 2004, **24**(3):945-953.
54. Zahn K, Inui M, Yukawa H: **Characterization of a separate small domain derived from the 5' end of 23S rRNA of an alpha-proteobacterium.** *Nucleic Acids Res* 1999, **27**(21):4241-4250.
55. Peyretailade E, Peyret P, Metenier G, Vivares CP, Prensier G: **The identification of rRNA maturation sites in the microsporidian *Encephalitozoon cuniculi* argues against the full excision of presumed ITS1 sequence.** *J Eukaryot Microbiol* 2001:60S-62S.
56. Katiyar SK, Visvesvara GS, Edlind TD: **Comparisons of ribosomal RNA sequences from amitochondrial protozoa: implications for processing, mRNA binding and paromomycin susceptibility.** *Gene* 1995, **152**(1):27-33.
57. Abeyratne PD, Nazar RN: **Parallels in rRNA processing: conserved features in the processing of the internal transcribed spacer I in the pre-rRNA from *Schizosaccharomyces pombe*.** *Biochemistry* 2005, **44**(51):16977-16987.
58. Kurland CG, Collins LJ, Penny D: **Genomics and the Irreducible Nature of Eukaryotic Cells.** *Science* 2006, **312**:1011-1014.
59. Welting TJ, Kikkert BJ, Van Venrooij WJ, Puijn GJ: **Differential association of protein subunits with the human RNase MRP and RNase P complexes.** *RNA* 2006, **12**:1373-1382.
60. Marquez SM, Harris JK, Kelley ST, Brown JW, Dawson SC, Roberts EC, Pace NR: **Structural implications of novel diversity in eucaryal RNase P RNA.** *RNA* 2005, **11**(5):739-751.
61. Collins LJ, Macke TJ, Penny D: **Searching for ncRNAs in eukaryotic genomes: Maximizing biological input with RNAmotif.** *Journal of Integrative Bioinformatics* 2004, **1**:61-77.
62. Missal K, Zhu X, Rose D, Deng W, Skogerbo G, Chen R, Stadler PF: **Prediction of structured non-coding RNAs in the genomes of the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*.** *J Exp Zool B Mol Dev Evol* 2006.
63. Saijou E, Fujiwara T, Suzuki T, Inoue K, Sakamoto H: **RBD-I, a nucleolar RNA-binding protein, is essential for *Caenorhabditis elegans* early development through 18S ribosomal RNA processing.** *Nucleic Acids Res* 2004, **32**(3):1028-1036.
64. Xin DD, Wen JF, He D, Lu SQ: **Identification of a *Giardia krr1* homolog gene and the secondarily anucleolate condition of *Giardia lamblia*.** *Mol Biol Evol* 2005, **22**(3):391-394.
65. Collins LJ, Poole AM, Penny D: **Using ancestral sequences to uncover potential gene homologues.** *Appl Bioinformatics* 2003, **2**(3 Suppl):S85-95.
66. Salinas K, Wierzbicki S, Zhou L, Schmitt ME: **Characterization and purification of *Saccharomyces cerevisiae* RNase MRP reveals a new unique protein component.** *J Biol Chem* 2005, **280**(12):11352-11360.

67. Embley TM, Martin W: **Eukaryote evolution, changes and challenges.** *Nature* 2006, **440**:623.
68. **NCBI HomePage** [<http://www.ncbi.nlm.nih.gov>]
69. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer S, Tacker M, Schuster P: **Fast folding and comparison of RNA secondary structures.** *Monatsh Chem* 1994, **125**:167-188.
70. Zuker M: **Calculating nucleic acid secondary structure.** *Curr Opin Struct Biol* 2000, **10**(3):303-310.
71. Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R: **RNAmotif, an RNA secondary structure definition and search algorithm.** *Nucleic Acids Res* 2001, **29**(22):4724-4735.
72. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25**(24):4876-4882.
73. Morgenstern B: **DIALIGN: multiple DNA and protein sequence alignment at BiBiServ.** *Nucleic Acids Res* 2004:W33-36.
74. Hofacker IL: **Vienna RNA secondary structure server.** *Nucleic Acids Res* 2003, **31**(13):3429-3431.
75. Höchsmann M, Toller T, Giegerich R, Kurtz S: **Local Similarity in RNA Secondary Structures.** *Proceedings of the IEEE Bioinformatics Conference(CSB 2003)* 2003:159-168.
76. Steffen P, Voß B, Rehmsmeier M, Reeder J, Giegerich R: **RNAshapes: an integrated RNA analysis package based on abstract shapes.** *Bioinformatics* 2006, **22**(4):.
77. Reeder J, Giegerich R: **Consensus Shapes: An Alternative to the Sankoff Algorithm for RNA Consensus Structure Prediction.** *Bioinformatics* 2005, **21**(17):3516-3523.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

